

STAT100 Problem Set 7: Confidence Intervals

You need to submit a Word document or PDF for this assignment. Make sure you do the following:

1. **Upload only one file/document in ELMS for Problem Set 7**
2. Include your name in the document in the upper left-hand corner. Under your name, write STAT 100 and your section number. Write Problem Set 7 centered on the page.
3. Number and letter your answers to the questions accordingly.
4. Carefully read all problems and follow all instructions.
5. Upload the assignment in ELMS before the deadline of Sunday 4/17 at 11:59 PM otherwise it is considered late. Make sure you save your document on your computer or email it to yourself so that you keep an electronic copy.

Students should refer to the Tutorial for Problem Set 7 as they are working on this problem set. All Tutorials for Problem Sets can be found in the STAT100 ELMS course under Modules.

For this assignment you need to use the *BodyFatPercentage.RData* file, which includes data collected for a study on obesity in adolescent girls. Below is a brief summary of the data set:

A medical researcher is studying obesity in adolescent girls. The researcher collected body mass index (BMI), body fat percentage, and other personal variables from 92 adolescent girls. The data collected by the researcher is a random sample representative of the population of all adolescent girls. The following variables were collected:

- BMI: the body mass index of each girl
- Pct_Fat: the body fat percentage of each girl
- Activity: the activity level for each girl (high, medium, or low)
- Menarche: whether the girl has achieved menarche (Yes or No)

We do not know the population standard deviation for adolescent girls' BMI or body fat percentage.

Problem Set 7 has three questions worth 25 points. Read each question carefully and follow all instructions. **Please follow these instructions for providing your responses:**

- For #1.a. and #2.a., you need to provide the R code that you use to generate output and provide the output displayed in the RStudio Console. In addition, you **MUST** include your name in the comments for the R code for #1.a. and #2.a.
- For all other questions on Problem Set 7, you should type your responses directly in the document you submit for Problem Set 7. You should **NOT** provide R code or RStudio output for those questions. If requested, be sure to *show all work to receive full credit*.

In Problem Set 7, for the first two questions you need to need to use sample data for a variable in the *BodyFatPercentage* data set to compute and interpret confidence intervals. For the third question, you need to solve problems involving confidence intervals.

1. (12 points) You will use sample data for the body fat percentage (*Pct_Fat*) variable from the *BodyFatPercentage* data file to compute and interpret confidence intervals.
 - a. Using RStudio, generate output showing the sample mean, sample standard deviation, and sample size for the *Pct_Fat* variable. *Refer to the Tutorial for Problem Set 7 for guidance on generating the appropriate output.* In the document you upload for this assignment, for 1.a. include an image showing the R code you used, including comments, and with the output from the RStudio Console. **IMPORTANT INSTRUCTIONS: To receive credit, you MUST include your name in the comments for the R code for #1.a.**
 - b. Based on the provided information about the population standard deviation and using the data from 1.a., compute a 95% confidence interval for the population mean body fat percentage of adolescent girls. Note that if the sample size is sufficiently large, you can use a z^* confidence multiplier instead of a t^* confidence multiplier. **DO NOT USE RStudio to compute this confidence interval;** do the calculations by hand and show all work in the document you upload for the assignment. *Refer to the Tutorial for Problem Set 7 for guidance and examples of showing work.* Round any intermediate values you use to three decimal places, and round the values in your confidence interval to three decimal places.
 - c. In your 95% confidence interval, what is the point estimate for the population mean body fat percentage of adolescent girls? *Provide a numeric value, rounded to three decimal places.*
 - d. In your 95% confidence interval, what is the margin of error? *Provide a numeric value, rounded to three decimal places.*
 - e. Interpret your 95% confidence interval for the population mean body fat percentage of adolescent girls. *Refer to the Tutorial for Problem Set 7 for guidance on interpreting confidence intervals.*
 - f. Based on the provided information about the population standard deviation and using the data from 1.a., compute a 99% confidence interval for the population mean body fat percentage of adolescent girls. Note that if the sample size is sufficiently large, you can use a z^* confidence multiplier instead of a t^* confidence multiplier. **DO NOT USE RStudio to compute this confidence interval;** do the calculations by hand and show all work in the document you upload for the assignment. *Refer to the Tutorial for Problem Set 7 for guidance and examples of showing work.* Round any intermediate values you use to three decimal places, and round the values in your confidence interval to three decimal places.
 - g. In your 99% confidence interval, what is the point estimate for the population mean body fat percentage of adolescent girls? *Provide a numeric value, rounded to three decimal places.*
 - h. In your 99% confidence interval, what is the margin of error? *Provide a numeric value, rounded to three decimal places.*

2. (7 points) You will use sample data for the activity level (*Activity*) variable from the *BodyFatPercentage* data file to compute and interpret a confidence interval.
- Using RStudio, generate two tables for the *Activity* variable: one table showing frequencies and one table showing percentages. *Refer to the Tutorial for Problem Set 7 for guidance on generating the tables.* In the document you upload for this assignment, for 2.a. include an image showing the R code you used, including comments, and with the output from the RStudio Console. **IMPORTANT INSTRUCTIONS: To receive credit, you MUST include your name in the comments for the R code for #2.a.**
 - Using the data from 2.a., compute a 90% confidence interval for the population proportion of adolescent girls with a “medium” activity level. **DO NOT USE RStudio to compute this confidence interval;** do the calculations by hand and show all work in the document you upload for the assignment. *Refer to the Tutorial for Problem Set 7 for guidance and examples of showing work.* Round any intermediate values you use to three decimal places, and round the values in your confidence interval to three decimal places.
 - In your 90% confidence interval, what is the point estimate for the population proportion of adolescent girls with a “medium” activity level? *Provide a numeric value, rounded to three decimal places.*
 - In your 90% confidence interval, what is the margin of error? *Provide a numeric value, rounded to three decimal places.*
 - Interpret your 90% confidence interval for the population proportion of adolescent girls with a “medium” activity level. *Refer to the Tutorial for Problem Set 7 for guidance on interpreting confidence intervals.*

3. (6 points) Answer the following questions about confidence intervals. **DO NOT USE RStudio for #3**; do the calculations by hand and show all work in the document you upload for the assignment. *Refer to practice problems in recent lecture slides for guidance.*
- In a recent study of 100 randomly selected statistics students, students were asked to provide the number of hours per week they spend studying for their statistics class. The results were used to compute confidence intervals for the population mean hours per week spent studying for statistics. The 99% confidence interval for the population mean hours per week that students spend studying for statistics was (6.85, 7.49). In this confidence interval, what is the sample mean (\bar{X}) hours per week spent studying for statistics? Your answer should be a numeric value rounded to two decimal places. *Show all work and explain how you determined your answer.*
 - Based on data from the same study, a 90% confidence interval for the population mean hours per week spent studying for statistics was also calculated. The lower endpoint of the 90% confidence interval was 6.96 and the margin of error (m) was 0.21. In this confidence interval, what is the upper endpoint? Your answer should be a numeric value rounded to two decimal places. *Show all work and explain how you determined your answer.*
 - If we want to create a 95% confidence interval for a population proportion with a margin of error (m) of 0.0225, what sample size is necessary? Use the simplified formula ($n = \frac{1}{m^2}$). *Show all work and round your final answer up to the next largest whole number.*